Identifying Residual Structure in Intrinsically Disordered Systems

Lessing, Joshua; Roy, Santanu; Reppert, Mike; Baer, Marcel; Marx, Dominik; Jansen, Thomas La Cour; Knoester, Jasper; Tokmakoff, Andrei

Published in:
Journal of the American Chemical Society

DOI:
10.1021/ja2114135

Identifying residual structure in intrinsically disordered systems: A 2D IR spectroscopic study of the GVGXPGVG peptide.

Joshua Lessing[1], Santanu Roy[2], Mike Reppert[1], Marcel Baer[3†], Dominik Marx[3], Thomas La Cour Jansen[2], Jasper Knoester[2], Andrei Tokmakoff[1]*

[1] Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139 USA

[2] Center for Theoretical Physics and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

[3] Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany

Supporting Information
Methods

Peptides

The peptides were prepared using Fmoc-based solid phase peptide synthesis performed on an automated peptide synthesizer. The peptides were then purified by reversed-phase HPLC, desalted and subsequently dissolved in a HCl solution and lyophilized 3 times in order to remove TFA[1,2].

2D IR Spectroscopy

To perform these experiments samples for FTIR and 2D IR spectroscopic measurements were first dissolved in D_2O and lyophilized in order to exchange the acidic hydrogen's of the peptide for deuterium. The resulting dry deuterated peptide was then dissolved in a pH = 1.0 DCl in D_2O solution at a concentration of 18.2 mg/ml. No aggregation was observed in the 2D IR spectra at this concentration, in addition for the Val peptide no spectral changes were observed over a range of 10 to 50 mg/ml. Next the sample was then placed between two 1mm thick CaF_2 windows separated by a 50μm Teflon spacer creating a ~25 μL sample volume. The FTIR spectra were collected on a dry air purged Nicolet 380 FTIR spectrometer at a spectral resolution of 2.0 cm^{-1} and averaged for 64 one-second scans. The 2D IR spectra were collected as described previously using 100 fs broad band mid-infrared pulses and a ZZZZ polarization geometry[3]. This data was collected using a waiting time of τ_2 = 150 fs and a τ_1 evolution time that was scanned to 3.1 ps for the rephasing spectrum and 2.5 ps for the non-rephasing spectrum in 4 fs time steps. All experiments were performed at pH = 1.0 in order to eliminate the $-COO^-$ stretch at 1590 cm^{-1} by deuterating the C-terminus creating the –COOD stretching vibration at 1722 cm^{-1}. In the case of figure S10 the 2D IR data was collected in a ZZYY polarization geometry at both pH = 1.0 and 7.0 for reference with no other alterations to experimental conditions.

MD Simulations

Local site energies as well as FTIR and 2D IR spectra were simulated for conformers obtained from molecular dynamics trajectories using a structure-based spectroscopic model. Eleven initial configurations have been extracted from a 200 nanosecond trajectory using the OPLS force field at 300 K (see Ref. 5 for technical details). The selected structures are representative of the main structural and dynamical motifs as found in extensive simulations carried out over a wide range of temperatures[4-7]. They contain representative structures with and without internal peptide-peptide hydrogen bonds in the extended and collapsed state. Those with internal hydrogen bonds provide distinct samples for different hydrogen bonding patterns close to the turn which dominate the hydrogen bonded stabilized structures in the trans-proline simulations[5,6]. This choice of initial conformations enables us to carry out rather short simulations compared to the relaxation times of the peptide which is on the order of nanoseconds[4,5] focusing on the local and short relaxations (picosecond time scale) which are probed in the IR experiment. These eleven configurations were then mutated to obtain the initial configurations of Ala and Gly. Next, MD simulations in NVT ensembles at 300 K were performed on these conformers using Gromacs-4.0.7 and an all atoms OPLS force field for the peptide which was solvated in a box of TIP3P water[4-6]. D_2O was used to reproduce the experimental conditions and all acidic protons of the peptides were deuterated. These simulations were carried out with 2 fs time steps and the LINCS algorithm was used to constrain all bond lengths[7]. For temperature coupling the Nose-Hoover thermostat was used[8,9]. The reaction field method was used to treat the long range electrostatic interactions with a cutoff of 1.4 nm. The length of each trajectory was 2.5 ns, where snapshots were saved every 10 fs yielding 2750000 conformations for each peptide studied.

Hydrogen bond criteria

A peptide-peptide hydrogen bond was defined as having an $O_{peptide}\cdots N_{peptide}–D_{peptide}$ angle ≤ 30° and a $O_{peptide}\cdots N_{peptide}$ distance ≤3.5Å and a peptide-water hydrogen bond was defined as having an $O_{peptide}\cdots O_{water}–D_{water}$ angle ≤30° and a $O_{peptide}\cdots O_{water}$ distance ≤ 3.5Å.

Binning of peptide structures

The hydrogen bonding criteria were used to define the structural bins for all peptides. The structural bins are categorized based on whether the carbonyl oxygen of the proline amide unit forms hydrogen bonds with the amide hydrogen ($HB_{peptide}$) or with the hydrogen of the solvent ($HB_{solvent}$). This definition created six structural bins (2/0, 1/0, 0/0, 0/1, 0/2 and 1/1) where the first and the second index represent the number of $HB_{peptide}$ and $HB_{solvent}$, respectively. The description of the spectral calculations for these bins is given below.

Spectral Modeling

A vibrational Hamiltonian was constructed for every snapshot by employing a Hamiltonian of the form[10]:

$$H(t) = \sum_{i=1}^{N}\left[\omega_i(t)b_i^{\dagger}b_i - \frac{\Delta_i(t)}{2}b_i^{\dagger}b_i^{\dagger}b_ib_i\right] + \sum_{ij}^{N} J_{ij}(t)b_i^{\dagger}b_i + \sum_{i=1}^{N}\mu_i(t)\bullet E(t)\left[b_i^{\dagger}+b_i\right]$$

Here the summations run over the amide I sites in the peptide where for any amide site i, $b_i^{\dagger}$ and b_i are the Bosonic creation and annihilation operators for the amide I vibration,
$\omega_i(t)$ is the frequency, $\Delta_i(t)$ is the anharmonicity, $J_{ij}(t)$ is the coupling between the i^{th} and j^{th} site, $E(t)$ is the applied laser field and μ_i is the transition dipole.

Amide I site frequencies (ω_i) for secondary amide groups were calculated with a Stark effect based approach using the electric field and its gradient on the C, O, N and D atoms[11]. For the proline amide unit the field components were considered on the C, O, N and C_δ atoms[12]. Next, frequency shifts and couplings due to through bond effects with neighboring amide units were determined using Ramachandran angle based look up tables[12,13]. The non-nearest-neighbor couplings were then calculated using a transition charge coupling scheme[12,13]. Finally the anharmonicity needed for describing doubly excited states was set to a value of 16 cm^{-1} for all amide units[10]. We made a 14 cm^{-1} systematic red-shift of the frequencies of all amide units except proline, because the frequency maps developed for the secondary amide units overestimate their frequency by the same amount.[12]

The FTIR and 2D IR spectra were calculated by solving the time dependent Schrödinger equation. This was accomplished using a numerical integration of the Schrodinger equation scheme[14]. For this scheme the Hamiltonian was considered to be constant for the 10 fs time intervals between successive saved structures of a MD trajectory. This approximation reduces the problem to solving the time independent Schrodinger equation for each time step in a trajectory and then using these time ordered matrices to obtain the linear and nonlinear response functions from which the FTIR and 2D IR spectra are calculated. An *ad hoc* vibrational life time (1 ps) was used in the spectral simulations[15]. The total time interval for calculating a sample for an FTIR spectrum was 2.5 ps and for a 2D IR spectrum it was 5 ps (τ_1=2.5 ps, τ_2=0 ps, τ_3=2.5 ps, where τ_1, τ_2, and τ_3 are the time delays between the pulses). For obtaining a spectrum of a particular bin we checked for every sample whether the bin was found in the middle of next 5 ps used for the spectral simulations. If it was found the spectrum was calculated.

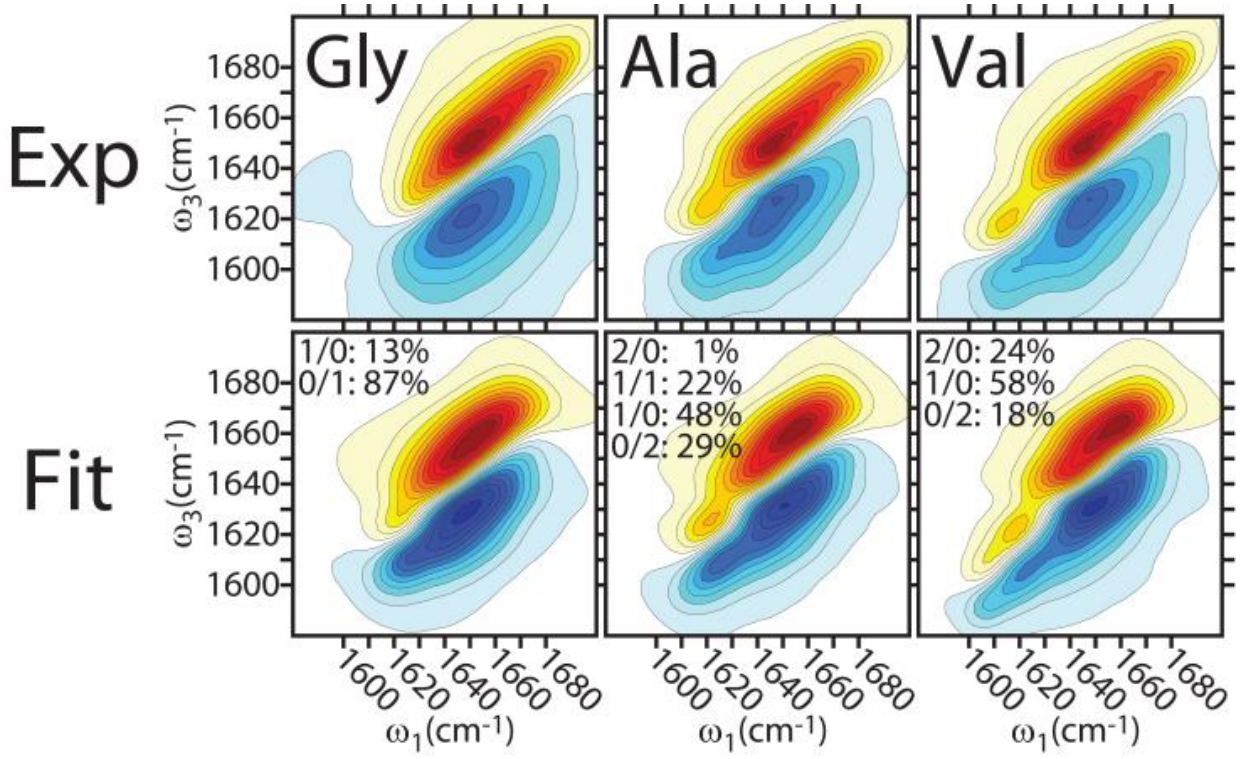

Figure (S1) - (Top row) Experimental 2DIR data, (Bottom row) - Fitted spectra, generated by fitting the experimental data over the frequency range 1575 to 1650 cm^{-1} with the simulated spectra for all six peptide bins using a genetic algorithm containing a least squares fitness function and no additional constraints. Since the amide I band maximum is located at ~1645cm-1 this fit range allowed for fitting both the proline peak shift and its intensity relative to the main amide band. Listed to the right of each surface is the population of individual HB bin's predicted from the fitting routine. Contours for all 2D IR surfaces presented in this paper are plotted from -1.0 to 1.0 in equally spaced steps of 0.08.

Bin	Population (%)		
	200 ns	27.5 ns	Fit
2/0	5.9	4.9	~25
1/1	3.2	1.3	<5
1/0	12.7	11.5	50-60
0/0	19.6	19.1	<5
0/1	47.1	49.9	<5
0/2	11.1	12.9	~20

Table (S1) – Bin populations calculated from a 200 ns MD trajectory, 27.5 ns MD trajectory, and the 2D IR experimental surface fitting for the Val peptide.

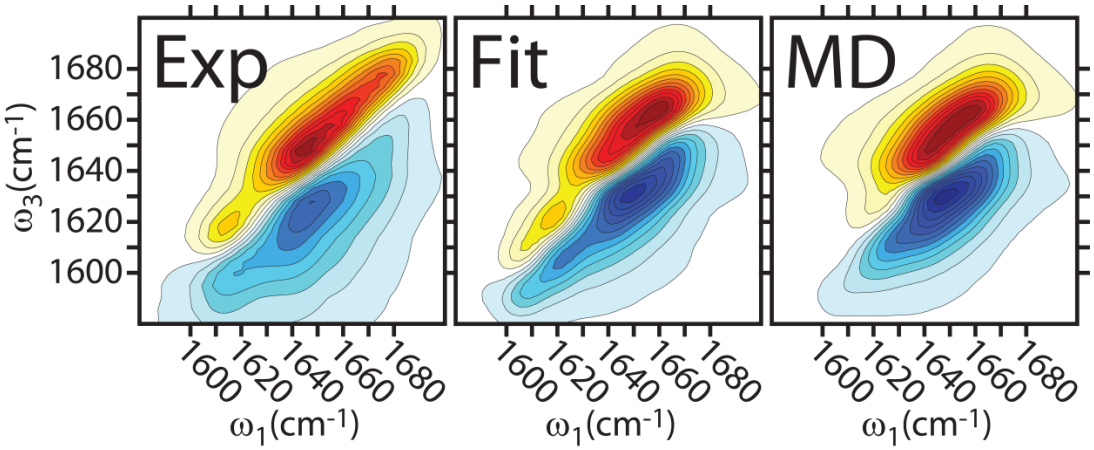

Figure (S2) – Comparison of the experimental spectrum for the Val peptide to simulated spectra using the bin populations generated for the fitting of the experimental data and populations calculated from a 200 ns MD trajectory.

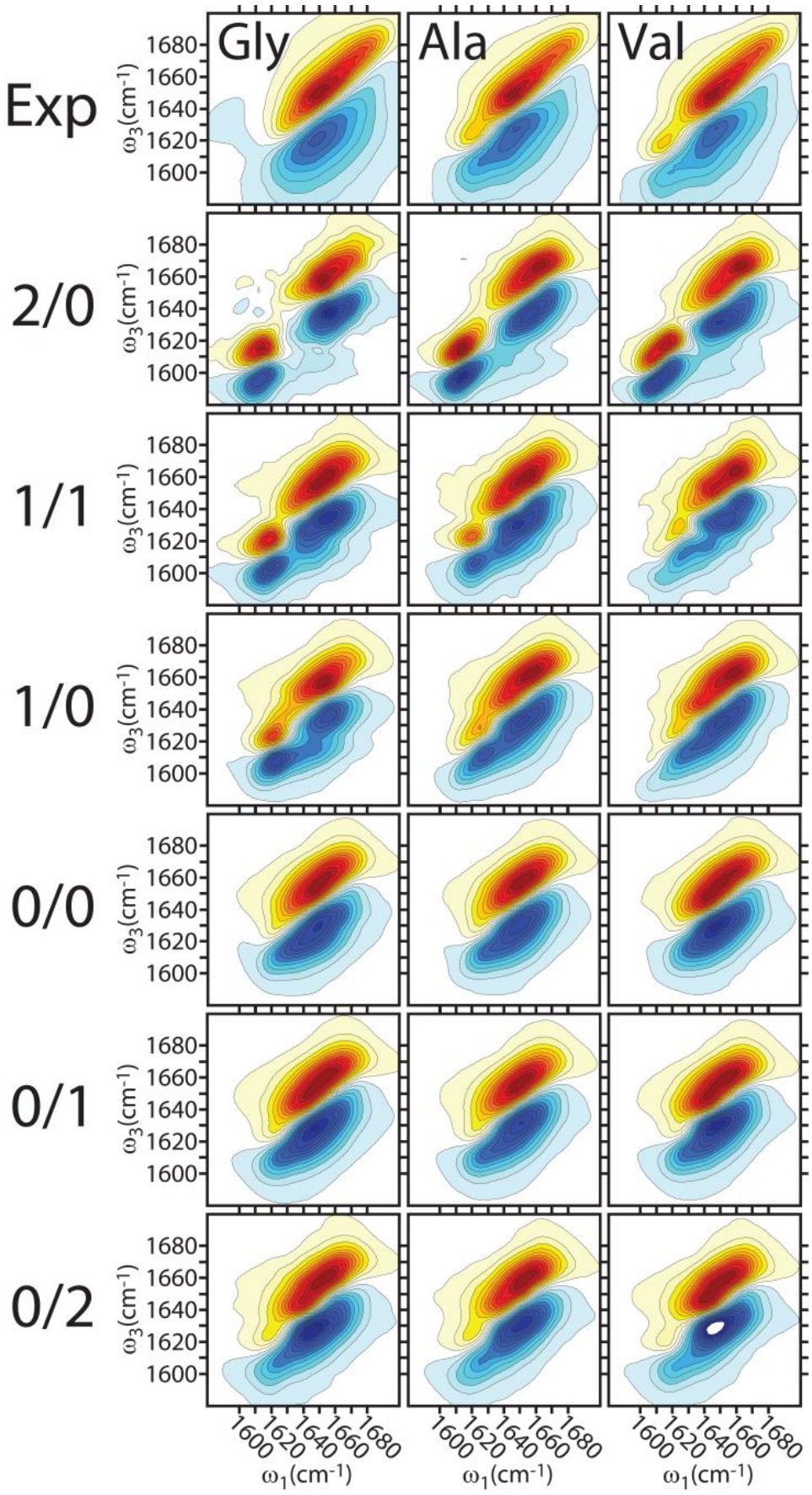

Figure (S3) - Comparison of the experimental data (Top row) to the six simulated surfaces generated for each peptide.

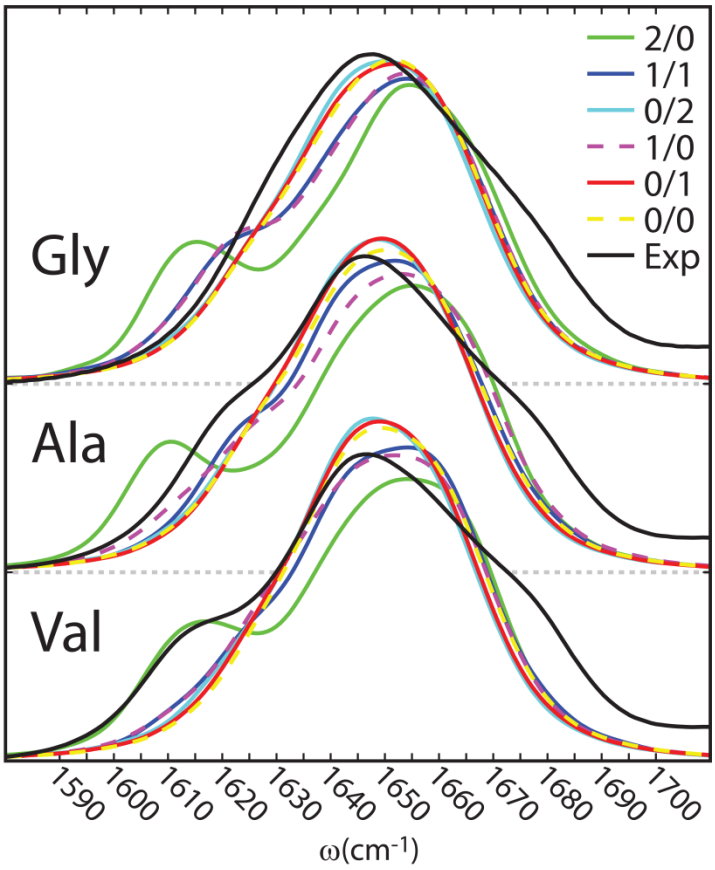

Figure (S4) – Comparison of experimental and simulated FTIR spectra for the Gly, Ala and Val peptides in the six hydrogen bond bins. Note the non-zero offset of the blue side of the spectrum is the result of a broad –COOD stretching vibration centered at 1722 cm^{-1}.

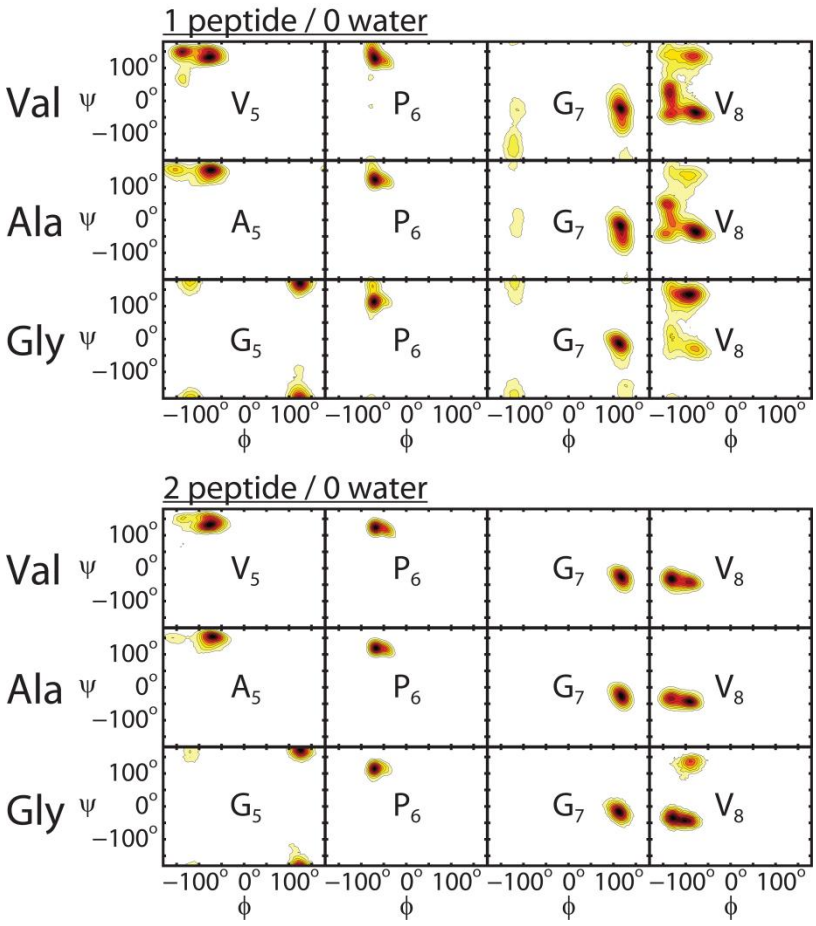

Figure (S5) - Ramachandran plots for the four central turn residues ($X_5P_6G_7V_8$) for the 1/0 and 2/0 hydrogen bond bins. Each plot was generated by culling structures based on hydrogen bonding criteria from 22 nsec of MD simulations for each of the three peptides GVGVPGVG, GVGAPGVG, and GVGGPGVG. Culling was performed using the g_hbond utility implemented in GROMACS 4.5 using an acceptor--donor--H cutoff angle of 30^o and a donor-acceptor cutoff distance of 3.5 Å. Ramachandran angles were likewise extracted using the GROMACS g_chi utility for each trajectory.

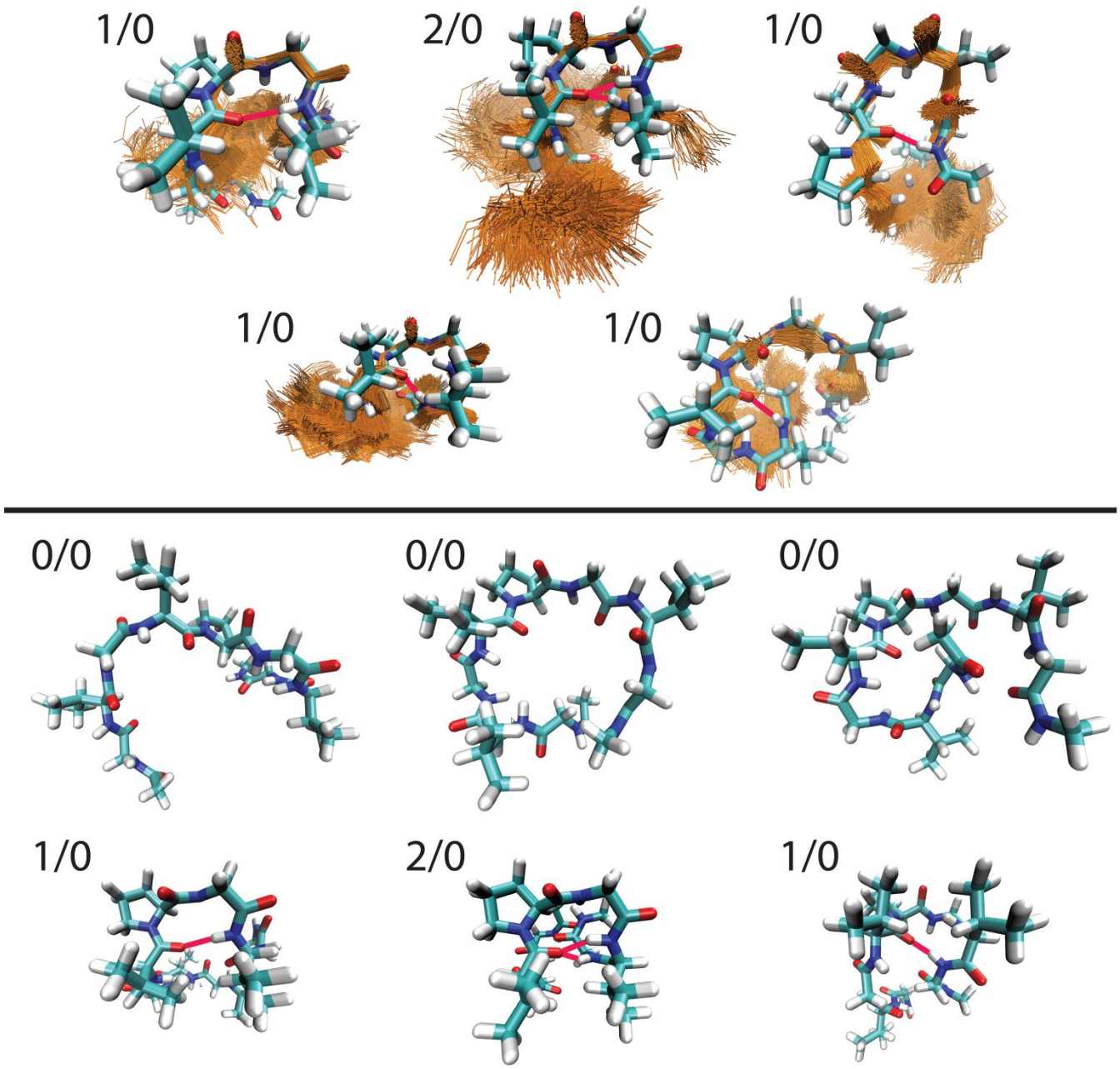

Figure (S6) - The images above show representative structures from the 1/0 and 2/0 hydrogen bonding bins selected from the Val MD trajectories used in this work. Hydrogen bonds are labeled for a single selected structure which is fully represented, with an ensemble of similar structures (peptide backbone only) is shown in gold for comparison.

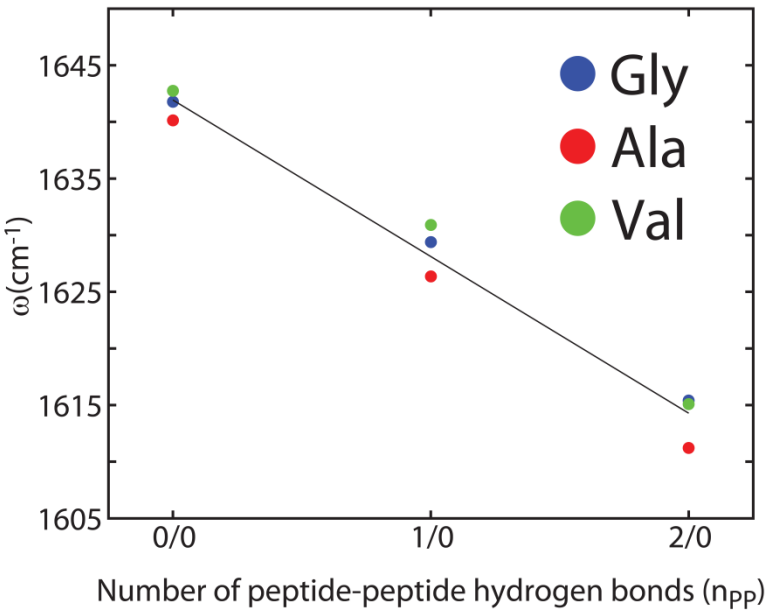

Figure (S7) - Average frequency of the simulated proline peak maximum for the 0/0, 1/0 and 2/0 bins. The linear fit to this data produced an equation relating the frequency of the proline peak and the number of peptide-to-proline hydrogen bonds: $\omega_p(cm^{-1}) \approx — 13.8 \cdot n_{PP} + 1642$.

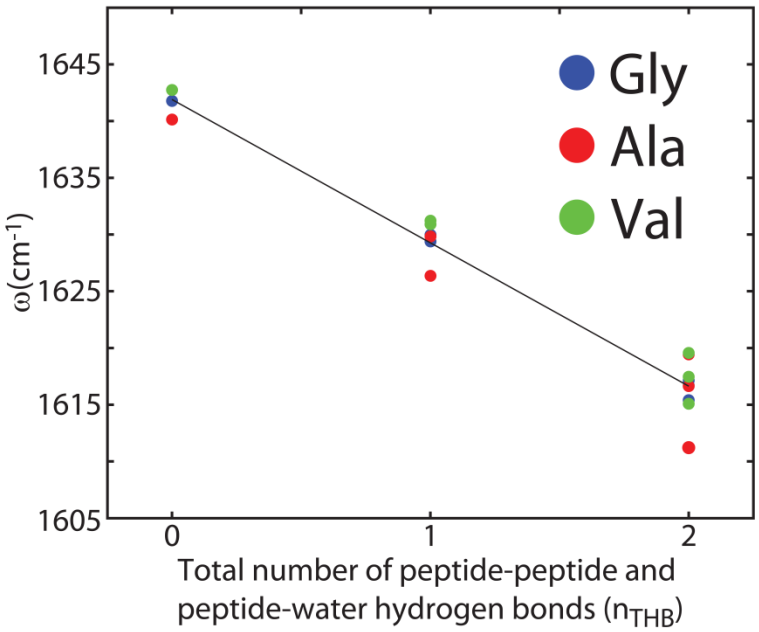

Figure (S8) - Average frequency of the simulated proline peak maximum for all six bins plotted as a function of the total number of peptide and water hydrogen bonds to proline. The linear fit to this data produced an equation relating the frequency of the proline peak and the total number of hydrogen bonds to proline: $\omega_p(cm^{-1}) \approx — 12.6 \cdot n_{THB} + 1642$.

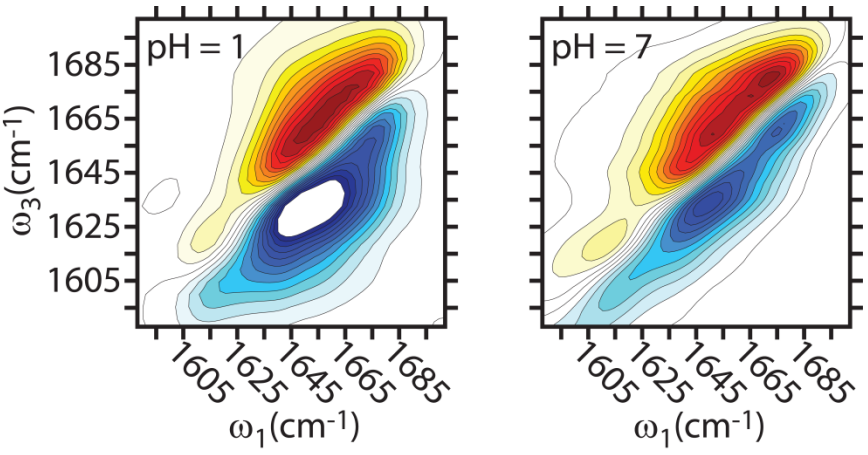

Figure (S9) – Val peptide 2D IR collected in ZZYY polarization at T = 10°C in a (Left) pH = 1.0, DCl in D_2O solution using τ_2 = 150 fs, and in a (Right) pH = 7.0, 50 mmol phosphate buffer solution using τ_2 = 0 fs. Changing the solvent conditions produces the appearance of on-diagonal broadening for the proline peak, a blue-shift of the main amide I band, and a reduction in the off-diagonal line width. The change in the appearance of the proline peak results from the generation of the COO moiety upon changing the pH, creating a peak at 1590 cm^{-1} from the COO asymmetric stretch. The blue shift of the main amide I band and the reduction in the off-diagonal line width observed at pH = 7 are likely due to the addition of buffer salts which produces a similar effect in pH = 1.0 samples.

References

(1) Andrushchenko, V. V.; Vogel, H. J.; Prenner, E. J. *J. Pept. Sci.* **2007**, *13*, 37-43.
(2) Roux, S.; Zékri, E.; Rousseau, B.; Paternostre, M.; Cintrat, J.-C.; Fay, N. *J. Pept. Sci.* **2008**, *14*, 354-359.
(3) Khalil, M.; Demirdöven, N.; Tokmakoff, A. *J. Phys. Chem. A* **2003**, *107*, 5258-5279.
(4) Hess, B.; Kutzner, C.; Spoel, D. v. d.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.
(5) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657-1666.
(6) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.
(7) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463-1472.
(8) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511-519.
(9) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695-1697.
(10) Hamm, P.; Lim, M.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *102*, 6123-6138.
(11) Jansen, T. L. C.; Knoester, J. *J. Chem. Phys.* **2006**, *124*, 044502.
(12) Roy, S.; Lessing, J.; Meisl, G.; Ganim, Z.; Tokmakoff, A.; Knoester, J.; Jansen, T. L. C. *J. Chem. Phys.* **2011**, *135*, 234507.
(13) Jansen, T. L. C.; Dijkstra, A. G.; Watson, T. M.; Hirst, J. D.; Knoester, J. *J. Chem. Phys.* **2006**, *125*, 044312.
(14) Jansen, T. L. C.; Knoester, J. *J. Phys. Chem. B* **2006**, *110*, 22910-22916.
(15) Kwac, K.; Lee, H.; Cho, M. *J. Chem. Phys.* **2004**, *120*, 1477-1490.